استناد به این مقاله: حریری، نجلا؛ طاهری، مهدی؛ فتاحی، رحمت‌الله؛ نوشین‌فرد، فاطمه (۱۳۹۲). میانکنش‌پذیری نظام‌های فراداده‌ای و موتورهای کاوش وب: تحولات و رویکردهای جاری. پژوهش‌نامه کتابداری و اطلاع‌رسانی،۳ (۲)،۲۵۲-۲۳۳.

میانکنش‌پذیری نظام‌های فراداده‌ای و موتورهای کاوش وب: تحولات و رویکردهای جاری

دکتر نجلا حریری[۱]؛ سید مهدی طاهری[۲]؛ دکتر رحمت الله فتاحی[۳]؛ دکتر فاطمه نوشین فرد[۴]

دریافت: ۹۰ ۱۱/۶ پذیرش: ۱۳۹۰/۱۲/۳

چکیده

هدف: *هدف از پژوهش حاضر تبیین مفاهیم، چالش‌ها، و رویکردهای جاری در حوزه میانکنش‌پذیری نظام‌های فراداده‌ای و موتورهای کاوش عمومی وب است.*

روش: *پژوهش حاضر، پژوهشی مفهومی است که با رویکرد تحلیلی-انتقادی، و با تاکید بر پژوهش‌های انجام شده به عنوان شواهد مرتبط، در راستای تحقق هدف پژوهش گام بر می‌دارد. بدین منظور، ضمن تقسیم‌بندی رویکردهای حوزه مورد مطالعه به چهار بخش اصلی، در هر بخش ابتدا بحثی مفهومی بر رویکرد مرتبط و دلایل پیدایش آن ارائه شده است، و سپس مروری تحلیلی و هدفمند بر پژوهش‌های مرتبط با آن رویکرد انجام می‌گیرد. در پایان هر بخش با دیدگاه انتقادی، چالش‌ها و روندهای جاری مرتبط با رویکرد مورد بحث، بررسی می‌گردند.*

یافته‌ها: *چهار رویکرد عمده در حوزه میانکنش‌پذیری نظام‌های فراداده‌ای و موتورهای کاوش عمومی وب شناسایی شدند. دلیل پیدایش این رویکردها، متاثر از کاربرد و آزمون اثر بخشی نظام‌های فراداده‌ای در افزایش سطح دسترس‌پذیری اشیای محتوایی وب از طریق موتورهای کاوش عمومی وب است. یافته‌ها حاکی از آن است که بستر نحوی پیاده‌سازی نظام‌های فراداده‌ای، بر کاربردپذیری آنها در بافت‌های اطلاعاتی مختلف، امکان پیاده‌سازی انواع طرح‌های فراداده‌ای متناسب با نیازهای جامعه فراهم کننده خدمات اطلاعاتی پیوسته، و افزایش نمایه‌پذیری، بازیابی‌پذیری، پیدانمایی، و رتبه‌بندی اشیای محتوایی وب در موتورهای کاوش عمومی وب بسیار تاثیر گذار است. موتورهای کاوش عمومی وب به دلایل مختلف، نسبت به برچسب‌ها و فرابرچسب‌های زبان نشانه‌گذاری فرامتن (HTML) واکنش نشان می‌دهند. با این وجود، محدودیت تعداد برچسب‌ها و فرابرچسب‌ها، و کارکرد صرفا نمایشی این زبان، کارایی این زبان را در سازماندهی اشیای محتوایی وب، و قابلیت پیاده‌سازی*

۱. دانشیار گروه کتابداری و اطلاع رسانی دانشگاه آزاد اسلامی، واحد علوم و تحقیقات تهران nadjlahariri@hotmail.com

۲. دانشجوی دکترای علوم کتابداری و اطلاع‌رسانی دانشگاه آزاد اسلامی، واحد علوم و تحقیقات تهران

۳. استاد گروه کتابداری و اطلاع‌رسانی دانشگاه فردوسی مشهد

۴. استادیار گروه کتابداری و اطلاع‌رسانی دانشگاه آزاد اسلامی، واحد علوم و تحقیقات تهران

دیگر طرح‌های فراداده‌ای استاندارد را کاهش می‌دهد. اما زبان نشانه‌گذاری گسترش‌پذیر (XML) با ارائه بستری قابل توسعه، علاوه بر فراهم نمودن امکان پیاده‌سازی تمامی طرح‌های فراداده‌ای استاندارد، نمایه‌پذیری و بازیابی‌پذیری اشیای محتوایی وب را به سطح کامل می‌رساند. بنابراین نظام‌های فراداده‌ای مبتنی بر این زبان، کارآمدترین ابزار برای دسترس‌پذیر نمودن اشیای محتوایی وب بوسیله موتورهای کاوش عمومی هستند.

کلیدواژه‌ها: *نظام‌های فراداده‌ای، موتورهای کاوش وب، میانکنش‌پذیری، بستر نحوی پیاده‌سازی*

مقدمه و بیان مسئله

دسترسی به اطلاعات و اشیای محتوایی وب نیز مانند دیگر محیط‌های اطلاعاتی بدون سازماندهی آنها دشوار است. فراداده، پاسخ نظام‌های سازماندهی دانش به محیط اطلاعاتی وب است. گسترش طرح‌های فراداده‌ای متعدد همزمان با تعمیم وب در اواسط دهه ۱۹۹۰ در همین راستا بوده است. استفاده از فراداده بیشتر در محیط‌های کتابخانه‌ای و پایگاه‌های اطلاعاتی انجام می‌گیرد. بنابراین برای دسترسی به پیشینه‌های فراداده‌ای – و به پیروی از آن اشیای محتوای مرتبط– مراجعه به فهرست‌ها، و نظام‌های جستجو و بازیابی اطلاعات محیط‌های یاد شده به صورت مجزا ضروری است. از سوی دیگر، از اواخر دهه ۱۹۹۰ ابزارهایی توسعه یافته‌اند که محتوای وب را به صورت خودکار سازماندهی، و جستجوپذیر می‌نمایند. این ابزارها – بویژه موتورهای کاوش عمومی– از محبوبیت بیشتری نزد کاربران وب برخوردارند. به نحوی که برخی پژوهش‌ها حاکی از آن است که ۸۸٪ زمانی که کاربران صرف جستجو در وب می‌کنند، در محیط موتورهای یاد شده است (بیفت[1] و کاستیلو[2]، ۲۰۰۵). از این رو، دسترس‌پذیر نمودن پیشینه‌های فراداده‌ای از طریق ابزارهای کاوش وب، همواره مد نظر طراحان و پیاده کنندگان نظام‌های فراداده‌ای بوده است. توجه به این مهم، از سوی نظام‌های کاوش وب نیز، میانکنش‌پذیری[3] هر دو نظام را در پی داشته است (طاهری[4] و حریری[5]، ۲۰۱۲). امکان دسترسی به پیشینه‌های فراداده‌ای از طریق موتورهای کاوش عمومی،

---

[1]. Bifet
[2]. Castillo

۳. اصطلاح "میانکنش‌پذیری" دارای مفهوم گسترده‌ای است، و به قابلیت کار متقابل میان دو یا چند نظام (سیستم)، محصول، سازمان، و غیره به منظور تبادل داده‌ها و خدمات، انجام وظایف مشترک، و جز آن اشاره دارد. در این پژوهش منظور از میانکنش‌پذیری، تمهیدات و اقداماتی است که نظام‌های فراداده‌ای و موتورهای کاوش برای قابلیت کار متقابل با یکدیگر انجام داده‌اند. این مفهوم از " قابلیت موتور جستجو در نمایه سازی عناصر فراداده‌ای" گسترده تر است و به تلاش‌های هم بافت فراداده‌ای و هم بافت موتورهای کاوش اشاره دارد. به عنوان مثال، انتخاب بستر نحوی XML به عنوان مبنای پیاده سازی و ذخیره سازی از سوی نظام‌های فراداده ای، و نمایه سازی اشیای (مدارک) مبتنی بر XML توسط روبات‌های موتورهای کاوش از مصادیق میانکنش‌پذیری آنهاست. یعنی اقداماتی که به تعامل بیشتر آن دو با هدف افزایش دسترس پذیری اطلاعات کتابشناختی (عناصر فراداده‌ای) می‌انجامد.

[4]. Taheri
[5]. Hariri

علاوه بر افزایش سطح دسترس‌پذیری اشیای محتوایی مرتبط، جستجوی یکپارچه را در بر خواهد داشت. در این صورت، موتورهای کاوش همانند یک درگاه[1] برای نظام‌های اطلاعاتی مختلف وب عمل خواهند نمود و موجب تولید ارزش افزوده می‌شوند.

در سال‌های اخیر، حوزه فراداده مطالعات گسترده‌ای را به خود اختصاص داده است. اما بیشتر این مطالعات از لحاظ ماهوی جنبه توصیفی دارد، و بر ویژگی‌های استانداردهای فراداده‌ای و پیاده سازی آنها در نظام‌های اطلاعاتی متمرکز بوده‌اند. تنها بخش محدودی از آنها - با رویکرد پژوهشی- به اثر بخشی فراداده در بازیابی اشیای محتوایی وب پرداخته‌اند (محمد[2]، ۲۰۰۶). معدود پژوهش‌های انجام گرفته نیز بیشتر جنبه کاربردی داشته‌اند. و کمتر پژوهشی با رویکرد مفهومی، موضوع فراداده، و مفاهیم و رویکردهای مرتبط با آن، و تعامل و یکپارچگی آن با ابزارهای کشف[3] وب را مد نظر قرار داده است. مقاله حاضر در قالب پژوهشی مفهومی و با تاکید بر پژوهش‌های انجام شده به عنوان شواهدی مرتبط، به بررسی مفاهیم، چالش‌ها، و رویکردهای جدید حوزه میانکنش‌پذیری نظام‌های فراداده‌ای و موتورهای کاوش عمومی وب می‌پردازد.

با بررسی تحلیلی تلاش‌های پیشین و جاری مرتبط با حوزه میانکنش‌پذیری نظام‌های فراداده‌ای و موتورهای کاوش، می‌توان این تلاش‌ها را بر مبنای چهار رویکرد کلی تقسیم نمود: رویکرد نخست، تلاش‌هایی که از برچسب‌ها و فرا برچسب‌های زبان نشانه گذاری فرامتن (HTML) برای سازماندهی اشیای محتوایی (صفحات وب)، و افزایش بازیافت پذیری[4] و رتبه‌بندی آنها در موتورهای کاوش عمومی وب استفاده می‌کردند؛ در رویکرد دیگر، کاربرد عناصر فراداده‌ای طرح‌های استاندارد مد نظر قرار گرفته‌اند؛ رویکرد سوم، تلاش‌هایی که در راستای رویکردهای اول و دوم (رویکرد دو وجهی) بوده‌اند؛ یعنی به طور همزمان اثر بخشی فرا برچسب‌های زبان نشانه‌گذاری فرامتن و عناصر استانداردهای فراداده‌ای را مورد بررسی قرار داده‌اند. و در رویکرد چهارم، پیشینه‌های مبتنی بر طرح‌های فراداده‌ای استاندارد در بستر نحوی جدید پیاده سازی - زبان نشانه‌گذاری گسترش‌پذیر (XML)، مبنای توصیف و سازماندهی اشیای محتوایی وب و دسترس پذیر نمودن آنها از طریق موتورهای کاوش وب قرار گرفته‌اند.

روش‌پژوهش

1. Portal
2. Mohamed
3. Discovery tools
4. Retrievability

پژوهش حاضر یک پژوهش مفهومی[۱] است که به تبیین و بسط مفاهیم، و بررسی تحلیلی تحولات و رویکردهای مرتبط با حوزه میانکنش‌پذیری نظام‌های فرادادهای و موتورهای کاوش وب می‌پردازد. مطالب مقاله بر اساس رویکردهای شناسایی شده در حوزه مورد مطالعه، به چهار بخش تقسیم می گردد. در هر بخش، نخست، مفاهیم مرتبط تبیین و مرور می گردند. سپس پژوهش‌های مرتبط به عنوان شواهد و نمونه به صورت تحلیلی بررسی شده، و در پایان هر بخش، تحولات، چالش‌ها و رویکردهای مرتبط با دیدگاه انتقادی تحلیل می گردند. روش گردآوری داده‌ها، روش کتابخانه‌ای (اسنادی) است، و جامعه پژوهش، پژوهش‌های مرتبط انجام شده، و نیز روندها و تجربیات حوزه مورد مطالعه همانند دیگر پژوهش‌های مفهومی[۲] است.

برچسب‌ها و فرا برچسب‌های زبان نشانه‌گذاری فرامتن (HTML) و موتورهای کاوش

نخستین تلاش‌های انجام شده در حوزه میانکنش‌پذیری عناصر فرادادهای و موتورهای کاوش، به واکنش موتورهای کاوش به برچسب‌ها و فرابرچسب‌های زبان نشانه گذاری فرامتن (HTML) اختصاص داشت. زبان نشانه گذاری فرامتن قالبی است که داده‌ها را به منظور نمایش نشانهگذاری می کند، و توسط کنسرسیوم وب جهانی (W3C)[۳] طراحی شده است. کارکرد صرفا نمایشی این زبان موجب شده، تعداد معدودی برچسب ارائه دهد. این مسئله در مورد فرابرچسب‌ها با محدودیت بیشتری روبروست. استفاده از زبان یاد شده بیشتر برای اشیای محتوایی (صفحات وب) است که در محیط‌های ایستای[۴] وب منتشر می‌شوند. به عبارت دیگر، امکان درونداد آن در پایگاه‌های داده‌ای وجود ندارد، و فقط به عنوان قالبی برونداد (خروجی) برای نمایش اشیا به کار می‌رود. به دلیل توسعه دیگر زبان‌ها توسط کنسرسیوم وب –که از قابلیت‌های بیشتری برخوردارند (مانند زبان XML)– گسترش این زبان متناسب با تحولات حوزه وب،بسیار محدود بوده است[۵]. زبان نشانه گذاری فرامتن در دهه ۱۹۹۰، پر کاربردترین زبان نشانه گذاری

---

[1]. Conceptual Research (or Paper)

[۲]. پژوهش‌های (یا مقالات) مفهومی در حال حاضر جایگاه خاصی را در میان پژوهش‌ها به دست آورده اند، و برخی از مجلات بین المللی بخش ویژه‌ای از هر شماره خود را به این گونه پژوهش‌ها اختصاص می دهند. برای اطلاعات بیشتر نگاه کنید به این پیوند:

http://www.conceptualstudy.org/Conceptual%20Papers.htm

[3]. W3 consortium

[4]. Static environment

[۵]. لازم به ذکر است، نسل جدیدی از این زبان با نام XHTML مبتنی بر زبان نشانه گذاری گسترش پذیر (XML) تولید شده است که محدودیت‌های کمتری نسبت به نسل پیشین دارد.

داده‌های وب بود. از این رو، ابزارهای کاوش وب، که روزهای آغازین فعالیت خود را می‌گذراندند- تعامل بالایی با برچسب‌ها و فرابرچسب‌های این زبان یافتند. و برچسب‌ها و فرابرچسب‌های زبان HTML کاملا برای نرم افزارهای پویش و نمایه‌سازی موتورهای کاوش شناخته شده‌اند. در میان پژوهش‌های مرتبط با حوزه فراداده، پژوهش‌های اندکی بر مبنای این رویکرد طرح‌ریزی شدند. اثربخشی برچسب‌های "عنوان[۱]" و "متن[۲]"، فرابرچسب‌های "کلیدواژه[۳]" و "توصیف[۴]"، تکرار کلیدواژه‌ها، و فراپیوندهای زبان نشانه‌گذاری فرامتن بر بازیافت پذیری و رتبه‌بندی اشیایی محتوایی در پژوهش‌های یاد شده مد نظر قرار گرفت. مهم‌ترین این پژوهش‌ها مربوط به ترنر[۵] و برکبیل[۶] (۱۹۹۸)، کوئودو-تورورو[۷]، (۲۰۰۴)، و ژانگ و دیمیتروف (۲۰۰۵a) است.

ترنر و برکبیل (۱۹۹۸) اثر بخشی استفاده از فرا برچسب‌های زبان نشانه‌گذاری فرامتن (HTML) در بهبود بازیابی اشیای محتوایی وب از طریق موتورهای کاوش عمومی را با روش تجربی مورد پژوهش قرار دادند. ۲۰ شیء محتوایی در ۵ حوزه فرعی موضوعی از حوزه کشاورزی به عنوان جامعه پژوهش ایجاد شدند. ۴ شیء بدون فرابرچسب، ۴ شیء دارای فرابرچسب "کلیدواژه"، ۴ شیء با فرابرچسب "توصیف"، و ۴ شیء حاوی فرابرچسب‌های "کلیدواژه" و "توصیف" به صورت مشترک. سپس به جستجوی واژگان مشترک هر شیء، و فرابرچسب‌های آنها در موتورهای کاوش آلتاویستا[۸] و اینفوسیک[۹] پرداخته شد. بیانگر آن بود، استفاده از فرابرچسب "کلیدواژه" دسترس‌پذیری اشیا را افزایش می‌دهد، اما کاربرد فرابرچسب "توصیف" به تنهایی موثر نیست. بنابراین ایشان بکارگیری فرابرچسب "کلیدواژه" را به تولید کنندگان اشیای مبتنی بر زبان HTML توصیه نموده، و نمایه‌سازی این فرابرچسب‌ها را به موتورهای کاوش، به منظور بهبود بازیابی‌پذیری اشیاء پیشنهاد کردند.

پژوهش دیگر که توسط کوئودو- توررو (۲۰۰۴) به عنوان رساله دکترای وی با عنوان "بهبود بازیابی بوسیله داده کاوی برچسب‌های HTML برای کلیدواژه‌ها و کشف ساختار فراپیوندهای صفحات وب" انجام شد، به کشف جنبه‌های مختلف وب برای بهبود کیفیت نتایج بازیابی با استفاده از روش تجربی پرداخت. هدف اصلی این پژوهش بهبود کیفیت جستجوی وب با جستارهایی پیرامون تاثیرات درج

1. Title
2. Body
3. Keyword
4. Description
5. Turner
6. Brackbill
7. Quevedo-Torrero
8. Altavista
9. Infoseek

کلیدواژه در فرابرچسب‌های HTML [به عنوان فراداده]، و ساختارها و روابط فرامتن‌ها در بازیابی اطلاعات از شبکه جهانی بود. جامعه پژوهش وی را تعدادی از صفحات وب و نتایج چند موتور کاوش (گوگل، آلتاویستا، و...) برای یک سری از کلیدواژه‌های منتخب تشکیل می‌دادند. یافته‌های این پژوهش عبارت بودند از راهبردهایی برای بهبود رتبه‌بندی نتایج موتورهای کاوش (ویژگی‌های مدارک وب و راهبرد TAKER)؛ راهکارهایی برای سنجش شباهت صفحات (در قالب چارچوبی برای مقایسه صفحات و انتخاب صفحات مشابه بر اساس معیارهایی چند) و خوشه‌بندی صفحات وب مطابق با ساختار پیوندیشان؛ و فنونی برای داده کاوی[1] به منظور شناسایی روابط میان محل کلیدواژه‌ها و سلسه مراتب HTML. این پژوهش پیشنهادات مفیدی برای بهبود الگوریتم‌های نمایه‌سازی و رتبه‌بندی موتورهای کاوش ارائه نمود.

پژوهش ژانگ و دیمیتروف (۲۰۰۵a) به بررسی تاثیر مشخصات محتوایی صفحات وب بر پیدا نمایی[2] آنها در نتایج موتورهای کاوش پرداخت. جامعه پژوهش را تعدادی صفحه وب آزمایشی مشتق شده از یک صفحه عمومی "مرکز ملی مکمل و طب سنتی" – که به ۱۹ موتور کاوش معرفی شدند – تشکیل می‌دادند. پرسش اصلی پژوهش، "چگونه رتبه‌بندی یک سایت در نتایج یک موتور کاوش از دیدگاه توسعه دهندگان صفحات وب بهبود می‌یابد؟" بود. پژوهش با روش تجربی انجام گرفت. سه فن آماری (one-way ANOVA, two-way ANOVA, and T-test) برای تحلیل داده‌ها استفاده شدند. یافته‌-ها حاکی از آن بود که: تکرار کلیدواژه‌ها فقط در عنوان، تکرار کلیدواژه‌ها فقط در متن کامل، و تکرار کلیدواژه‌ها هم در عنوان و هم در متن کامل، باعث بهبود پیدانمایی صفحات وب در نتایج موتورهای کاوش می‌شوند. اما عواملی چون رنگ فونت، اندازه فونت، و ... تاثیری در بهبود پیدا نمایی ندارند.

نتایج این پژوهش‌ها واکنش مثبت موتورهای کاوش به اشیای محتوایی مبتنی بر زبان نشانه گذاری فرامتن (HTML) را تایید می‌کند. بنابراین میانکنش‌پذیری برچسب‌ها و فرابرچسب‌ها با موتورهای کاوش در سطح نسبتا مطلوبی وجود دارد. استفاده از فرابرچسب‌ها، کلیدواژه‌های خاص، تکرار کلیدواژه‌ها و نظیر آن از یک سو، و توجه روبات‌های موتورهای کاوش به این ویژگی‌ها از سوی دیگر، حاکی از تمهیدات هر دو سوی تعامل برای افزایش سطح دسترس پذیری است. بررسی نتایج این پژوهش‌ها، بیانگر نکاتی چند است. تمامی این پژوهش‌ها از روش تجربی استفاده نمودند. برخی راهکارهای به کار گرفته شده توسط پژوهش‌ها –از جمله تکرار کلیدواژه‌ها برای افزایش پیدا نمایی ژانگ و دیمیتروف (۲۰۰۵a)، در حال حاضر از سوی روبات‌های موتورهای کاوش نوعی تقلب محسوب می‌شود، و تاثیر خود را از دست داده

[1]. Data mining
[2]. Visibility

است (سرچ انجین توتوریال[1]، ۲۰۱۲). موتور کاوش آلتاویستا یکی از موتورهای کاوش اصلی مورد مطالعه بود، اما پس از خرید آن توسط شرکت یاهو، و توسعه موتور کاوش یاهو از سال ۲۰۰۴، موتور کاوش جدید جایگزین آلتاویستا در پژوهش‌های این حوزه شد.

محدودیت‌های زبان نشانه گذاری فرامتن (HTML) باعث شد، چندان مورد استقبال نظام‌های کتابخانه‌ای و پایگاه‌های اطلاعاتی وب که حاوی اطلاعات علمی و معتبر هستند، واقع نشوند، و تنها برای انتشار صفحات ثابت و عمومی این نظام‌ها مورد استفاده قرار گیرند. در همین راستا، فرا برچسب‌های این زبان نیز مناسب سازماندهی اشیای محتوایی معتبر وب ارزیابی نشدند و نتوانستند همانند طرح‌های فراداده‌ای استاندارد عمل کنند. از مهم‌ترین این محدودیت‌ها می‌توان به موارد ذیل اشاره نمود:

- محدودیت ذخیره‌سازی اشیای محتوایی در بستر زبان نشانه‌گذاری فرامتن، کاربردپذیری آن را در نظام‌های اطلاعاتی تمام متن –که بیشتر در محیط‌های پویا پیاده سازی می‌شوند– کاهش داده است. چرا که این قالب ماهیتا برای ذخیره‌سازی داده‌ها طراحی نشده بود. به عنوان مثال، امکان ذخیره‌سازی متن کامل شیئی مانند کتاب در یک فایل متبنی بر زبان HTML وجود ندارد. این ویژگی منجر به گرایش نظام‌ها به دیگر قالب‌های ذخیره سازی داده‌ها شد؛
- تعداد فرابرچسب‌های زبان نشانه‌گذاری فرامتن بسیار محدود است. بنابراین نقاط دسترسی اندکی را برای بازیابی اشیای محتوایی وب فراهم می‌نماید. همچنین برای توصیف اشیای محتوایی کارایی لازم را ندارد (طاهری و حریری، ۲۰۱۲؛ طاهری، ۱۳۸۷الف). به معنای دیگر، کارکردها و کاربردهای فراداده را در سطحی ضعیف پشتیبانی می‌کنند؛
- با وجود آن که موتورهای کاوش عمومی بر استفاده فرابرچسب‌ها با ارزش‌هایی (محتوایی) منحصر به فرد، برای افزایش نمایه‌پذیری صفحات (اشیای محتوایی)، و افزایش رتبه آنها در نتایج تاکید می‌کنند (گوگل[2]، ۲۰۱۲؛ یاهو، ۲۰۱۲)، اما در نظام جستجوی خود، امکان جستجوهای عنصر– پایه بر اساس فرابرچسب‌های زبان HTML را میسر نمی‌سازند؛
- فرابرچسب‌ها فقط بر اساس رویکرد جاسازی (درج)[3] در صفحات وب قابل استفاده هستد. بنابراین در محیط فهرست‌های رایانه‌ای و کتابخانه‌های دیجیتالی که مبتنی بر رویکرد پیوند[4] فراداده‌ها با اشیای محتوایی عمل می‌نمایند، کاربردی ندارند؛

---

1. Search Engine Tutorial
2. Google
3. Embedding
4. Linking or Attaching

- ارزش‌های[1] فرابرچسب‌ها مستند و استاندارد نیستند. یعنی از ابزارهای کنترل واژگان و فهرست‌های مستند برای تامین آنها استفاده نمی‌شود. بنابراین کاهش جامعیت در بازیابی را در پی خواهند داشت؛
- فرابرچسب‌ها بیشتر توسط پدیدآورندگان صفحات و یا منتشر کنندگان آنها که در حوزه سازماندهی اطلاعات و دانش غیر متخصص هستند، تولید می‌شوند. بدیهی است چنین پیشینه‌های فراداده‌ای غیر استاندارد بوده، قابلیت‌های پیشینه‌های فراداده‌ای استاندارد را که توسط متخصصان سازماندهی ایجاد می‌شوند، ندارند؛
- از آنجا که فرابرچسب‌ها از دید کاربران نهایی پنهان هستند، و روش‌های مشاهده آنها متداول نیست، شناسایی و انتخاب اشیای محتوایی به عنوان دو کارکرد از کارکردهای اصلی پیشینه‌های فراداده‌ای (کتابشناختی) از طریق آنها امکان پذیر نمی‌باشد.

پیشینه‌های فراداده‌ای مبتنی بر طرح‌های فراداده‌ای استاندارد و موتورهای کاوش

رویکرد دوم در حوزه میانکنش‌پذیری نظام‌های فراداده‌ای و موتورهای کاوش عمومی وب، استفاده از طرح‌های فراداده‌ای استاندارد برای سازماندهی اشیای محتوایی وب، و قرار دادن عناصر و پیشینه‌های مبتنی بر آن طرح‌ها در معرض نرم افزارهای خزنده- نمایه‌ساز[2] موتورهای کاوش بود. این طرح‌ها مجموعه‌ای از عناصر فراداده‌ای هستند که برای، توصیف، شناسایی، نگهداری، کشف و بازیابی، و ارتباط میان اشیای محتوایی به کار می‌روند (فتاحی و طاهری، ۱۳۸۸)، و توسط سازمان‌های معتبر بین‌المللی توسعه یافته‌اند. بخشی از این طرح‌ها پیش از توسعه وب (مانند قالب مارک)، و برخی بعد از تعمیم وب (مانند هسته دوبلین) گسترش یافتند. تفاوت‌های محیط وب نسبت به محیط‌های اطلاعاتی سنتی باعث شد، طرح‌هایی که بعد از تعمیم وب و متناسب با محیط جدید طراحی شدند، از استقبال بیشتری برخوردار شوند (طاهری، ۱۳۸۷ب). توجه به برخی از طرح‌های فراداده‌ای جدید از جمله طرح "فراداده‌ای هسته دوبلین (DCMI)" در اواخر دهه ۱۹۹۰ و اوایل دهه ۲۰۰۰ و در بافت محیط‌های اطلاعاتی الکترونیکی و شبکه‌ای به سرعت افزایش یافت.

پیشینه‌های مبتنی بر طرح‌های فراداده‌ای استاندارد، امکان پیاده‌سازی در بستر نحوی زبان‌های نشانه‌گذاری، و پایگاه‌های داده‌ای را دارند. اما از آنجا که گسترش طرح‌های فراداده‌ای جدید همزمان با

[1]. Value
[2]. Spider-indexer or Crawler-indexer

تعمیم وب، و استفاده گسترده از زبان نشانه‌گذاری فرامتن (HTML) بود، این زبان به عنوان بستر پیاده‌سازی طرح‌های فراداده‌ای انتخاب شد. واکنش مثبت روبات‌های موتورهای کاوش به این زبان نیز مزید علت بود. غنای عناصر طرح‌های فراداده‌ای، و ارتباط معناشناختی عناصر آنها با یکدیگر، گرایش فراهم کنندگان خدمات اطلاعاتی به جایگزینی این طرح‌ها با فرابرچسب‌های زبان نشانه‌گذاری فرامتن (HTML) را نشان می‌داد. با این وجود، محدودیت‌های زبان نشانه‌گذاری فرامتن امکان پیاده‌سازی همه طرح‌های فراداده‌ای (مانند مارک) را میسر نمی‌ساخت، اما برخی طرح‌ها (ازجمله هسته دوبلین) قابل پیاده‌سازی بودند. در همین راستا، برخی از پژوهش‌های حوزه فراداده به بررسی نمایه‌پذیری، بازیابی‌پذیری، و رتبه‌بندی اشیای محتوایی حاوی عناصر فراداده‌ای مبتنی بر طرح‌های استاندارد در قالب زبان نشانه‌گذاری فرامتن در موتورهای کاوش پرداختند. پژوهش‌های سوکویتن[1] (۲۰۰۰ وصفری[2] ۲۰۰۵)از آن جمله‌اند.

سوکویتن (۲۰۰۰) در پژوهش خود سودمندی طرح فراداده‌ای هسته دوبلین برای فرایند بازیابی را مورد مطالعه قرار داد. هدف اصلی این پژوهش تحلیل عینی و شاخص توانایی بازیابی عناصر کلیدی (عنوان، ناشر، پدیدآورنده، و موضوع) طرح فراداده‌ای هسته دوبلین جاسازی شده در صفحات وب بیست سازمان بزرگ دولتی و آموزشی استرالیا بود.

یافته‌های این پژوهش عبارت بودند از: عنصر عنوان هسته دوبلین به دلیل این که معمولا محتوای آن با محتوای برچسب عنوان زبان HTML صفحه وب مشابه است، کارآیی در فرایند بازیابی ندارد؛ عناصر پدیدآورنده، همکار، و ناشر می‌توانند برای بازیابی مفید باشند اما به دلیل عدم انسجام در قالب ثبت محتوا، ارزشی ندارند؛ عنصر موضوع اکنون به درستی استفاده نمی‌شود بنابراین ارزشی برای فرایند بازیابی ندارد. در این پژوهش پیشنهاداتی برای بهبود ایجاد پیشینه‌های فراداده‌ای ارائه شده است و رابطه بین طرح هسته دوبلین و "نظریه جستجوی اطلاعات[3]" کشف شد. در ایران نیز پژوهشی مشابه انجام شد؛ صفری (۲۰۰۵) در پژوهش خود به بررسی تاثیر عناصر فراداده‌ای هسته دوبلین بر کشف منابع وب و بهبود رتبه آنها در موتورهای کاوش پرداخت. جامعه پژوهش او را مقالات منتشر شده (۱۶ مقاله) بر روی وب سایت "مجله بین المللی ایرانی علم[4]" تشکیل می‌دادند. صفحات وب مربوط به مقالات به دو گروه آزمایش (شامل ۴ عنصر از ۱۵ عنصر طرح فراداده‌ای هسته دوبلین) و گروه گواه (بدون عناصر فراداده‌ای هسته

1. Sokvitne
2. Safari
3. Information foraging theory
4. Iranian international journal of science

دوبلین) تقسیم شده و واکنش موتورهای کاوش گوگل، لایکاس، آلتاویستا، و ... نسبت به آنها مورد ارزیابی قرار گرفت.

نتایج حاصل از این پژوهش – که با استفاده از آزمون یو مان-ویتنی[1] و به روش تجربی به دست آمد، عبارتند از: الف. تفاوت معناداری میان رتبه صفحاتی که از عناصر فراداده‌ای هسته دوبلین استفاده کردند و گروهی که استفاده نکردند (گروه گواه) وجود ندارد. بنابراین پاسخ سئوال اول منفی است. ب. تفاوت معناداری میان گروه گواه و آزمایش از لحاظ عملکرد بازیابی دیده نشد. به عبارت دیگر، عناصر فراداده‌ای هسته دوبلین تاثیر معناداری بر بازیابی صفحاتی که از این عناصر استفاده کردند نسبت به آنها که استفاده نکردند، نداشتند. پژوهش‌های مبتنی بر رویکرد دوم نشان دادند، استفاده از طرح فراداده‌ای هسته دوبلین در بستر زبان نشانه گذاری فرامتن (HTML) اثر بخشی چندانی در افزایش بازیافت پذیری و بهبود رتبه اشیای محتوایی در فهرست نتایج موتورهای کاوش ندارد. این دو پژوهش نیز همانند پژوهش‌های رویکرد اول از روش تجربی استفاده کردند. ویژگی دیگر آنها، انتخاب ۴ عنصر از عناصر هسته پانزده گانه (سطح ساده) طرح فراداده‌ای هسته دوبلین برای بررسی بود.

بهره‌مندی از عناصر فراداده‌ای طرح‌های استاندارد در قالب‌های زبان نشانه‌گذاری فرامتن (HTML)، علیرغم ارزش‌های افزوده بیشتری که نسبت به فرابرچسب‌های زبان نشانه‌گذاری فرامتن دارند، برای سازماندهی اشیای محتوایی، و افزایش دسترس‌پذیری آنها توسط ابزارهای کاوش عمومی مناسب نیستند. به عبارت دیگر، موتورهای کاوش فقط فرابرچسب‌های زبان نشانه‌گذاری فرامتن را مورد توجه قرار می‌دهند، و عناصر دیگر طرح‌ها برای آنها ارزش نمایه‌سازی ندارند. بنابراین می‌توان عدم نمایه‌سازی عناصر فراداده‌ای هسته دوبلین توسط موتورهای کاوش را، بستر نحوی پیاده‌سازی آنها دانست. از سوی دیگر، محدودیت‌های زبان نشانه‌گذاری فرامتن از لحاظ تعداد و ساختار برچسب‌ها و فرابرچسب‌ها، امکان تولید پیشینه‌های هسته دوبلین در بستر آن زبان به صورت مجزا را نیز فراهم نمی‌کند. و فقط می‌توان این پیشینه‌ها را بر اساس رویکرد جاسازی (درج) تولید نمود. این بدان معناست که پیاده‌سازی پیشینه‌های فراداده‌ای هسته دوبلین در بستر زبان نشانه گذاری فرامتن (HTML)، کاربردپذیری آنها در نظام‌های کتابخانه‌ای –به ویژه فهرست‌های رایانه‌ای- و پایگاه‌های اطلاعاتی را کاهش می‌دهد.

رویکرد دو وجهی (فرا برچسب‌های HTML و استانداردهای فراداده‌ای)

برای دستیابی به مزایای استفاده از طرح‌های فراداده‌ای استاندارد به دلیل غنی بودن مجموعه عناصر

---

[1]. Mann-whitney U

آنها، و روابط معناشناختی میان عناصر، و نیز فرابرچسب‌های زبان نشانه‌گذاری فرامتن (HTML) با قابلیت بالای نمایه‌پذیری و اثربخشی در رتبه‌بندی اشیای محتوایی، فراهم کنندگان خدمات اطلاعاتی پیوسته به پیاده‌سازی هر دو نظام تمایل نشان دادند. در زمان رشد این گرایش، اشیای محتوایی که در قالب زبان نشانه‌گذاری فرامتن تولید می‌شدند، هم دارای عناصر طرح‌های فرادادهای، و هم فرابرچسب‌های زبان نشانه‌گذاری فرامتن در بخش سرصفحه خود بودند. در این رویکرد نیز، تنها بستر نحوی مورد استفاده، زبان نشانه‌گذاری فرامتن بود.

بنابراین، همان طور که پیشتر اشاره شد، فقط امکان پیاده‌سازی برخی از طرح‌های استاندارد وجود داشت، و از بقیه طرح‌ها از جمله مارک –که بیشترین کاربرد را در بافت کتابخانه‌ای داشت- صرف نظر می‌شد. امکان پیاده‌سازی فرابرچسب‌های زبان نشانه‌گذاری فرامتن، و پیشینه‌های فرادادهای مبتنی بر طرح‌های استاندارد به صورت همزمان در صفحاتی که در قالب HTML ارائه شده بودند، انگیزه مقایسه دو نظام فرادادهای را بر انگیخت. بدین ترتیب، پژوهش‌هایی چند –که بیشترین تعداد را در حوزه میانکنش‌پذیری دارند- به بررسی تلاش‌های مبتنی بر رویکرد دو وجهی، و کارایی آن در سازماندهی و دسترس‌پذیری اشیای محتوایی وب اختصاص یافتند.

ارزیابی سودمندی فراداده‌ها به عنوان ابزاری برای افزایش بازیابی اطلاعات در یک مجموعه هفت‌گانه از موتورهای کاوش موضوع پژوهش هنشاو[1] و والاسکاس[2] (۲۰۰۱) بود. این پژوهش با روش تجربی بر روی صفحاتی منتخب از برخی شماره‌های مجله اینترنتی First Monday در دو گروه گواه (بدون فراداده) و گروه آزمون (شامل عناصر فرادادهای هسته دوبلین و فرابرچسب‌های توصیف و کلیدواژه HTML) انجام شد.

نتایج پژوهش نشان داد فراداده‌ها به تنهایی نقش مهمی را در افزایش احتمال نمایه‌سازی منابع و کسب رتبه بالا در نتایج موتورهای کاوش ایفا نمی‌کنند (هنشاو و والاسکاس، ۲۰۰۱). پس از وی، ژانگ و دیمیتروف (۲۰۰۴) عملکرد موتورهای کاوش مهم اینترنت (هفت موتور کاوش: گوگل، آلتاویستا، یاهو، لایکاس، و...) در ارتباط با دو گروه از صفحات وب (به عنوان جامعه پژوهش) – شامل عنصر موضوع از طرح فرادادهای هسته دوبلین و عنصر کلیدواژه از زبان HTML (گروه آزمون) همراه با تکرار کلیدواژه‌ها در عناصر یاد شده، و فاقد عناصر فرادادهای مذکور (گروه گواه) همراه با تکرار کلیدواژه‌ها در عنوان و متن کامل – را با روش تجربی مورد مطالعه قرار دادند. یافته‌ها حاکی از آن بود، تفاوت‌های

---

1. Henshaw, Robin
2. Valauskas

معناداری میان دو گروه گواه و آزمون وجود دارد، و بیشتر موتورهای کاوش مورد مطالعه (شش موتور) نسبت به حضور عناصر فراداده‌ای پاسخ مثبت، و برخی (یک موتور) پاسخ منفی دادند (ژانگ و دیمیتروف، ۲۰۰۴).

ایشان دوباره پژوهش دیگری پیرامون تاثیر پیاده سازی فراداده‌ها در یک صفحه وب بر روی عملکرد پیدایی آن صفحه در فهرست نتایج یک موتور کاوش انجام دادند. مطالعه بر روی ۴۰ صفحه وب آزمایشی دارای ویژگی‌هایی بر اساس جداول از قبل طراحی شده – که به ۱۹ موتور کاوش معرفی شده بودند – انجام می‌گرفت. ابتدا عوامل موثر داخلی (محل و تکرار کلیدواژه‌ها در یک عنصر فراداده‌ای) و خارجی (منبعی که کلیدواژه‌های مورد استفاده در عنصر فراداده‌ای از آن استخراج می‌شوند؛ مثلا عنوان صفحه) بر پیاده‌سازی فراداده‌ها شناسایی گردیدند. سپس با استفاده از روش تجربی، چگونگی تاثیر این عوامل بر پیدایی صفحات وب در فهرست نتایج یک موتور کاوش آزمایش شدند. نتایج بیانگر آن بود که فراداده مکانیزمی مناسب برای بهبود پیدایی صفحات وب است؛ منطقه (یا عنصر) موضوع فراداده نقشی مهم‌تری را نسبت به سایر مناطق یا عناصر فراداده‌ای بازی می‌کند و کلیدواژه‌های استخراج شده از خود صفحات، بویژه از عنوان و متن کامل، بسیار موثر می‌باشند (ژانگ و دیمیتروف، ۲۰۰۵b).

محمد (۲۰۰۶) پژوهش مشابهی را طرح‌ریزی کرد. وی تاثیر استفاده از فراداده در یافتن و رتبه‌بندی صفحات وب در موتورهای کاوش را بررسی نمود. در فاز اول، کاربردپذیری طرح‌های فراداده‌ای و تاثیر اشیای محتوایی که از لحاظ محتوایی همپوشانی دارند، مورد مطالعه قرار گرفت. و در فاز دوم، تاثیر افزودن فراداده بر بهبود رتبه صفحات وب مد نظر قرار گرفت. محمد اظهار داشت: فرابرچسب‌های زبان نشانه‌گذاری فرامتن (HTML) بیش از عناصر فراداده‌ای هسته دوبلین استفاده می‌شوند. و فراداده بر بهبود رتبه صفحات موثر است. آخرین پژوهش مربوط به شریف (۱۳۸۶) بود. هدف پژوهش او، تعیین میزان اثربخشی عناصر فراداده‌ای بر رتبه‌بندی صفحات وب توسط سه موتور کاوش عمومی "گوگل"، "یاهو"، و "ام‌اس‌ان" بود. این پژوهش با روش تجربی انجام شد، و شامل دو گروه گواه و آزمون بود. برای موتورهای کاوش اهمیت فرابرچسب‌های زبان نشانه‌گذاری فرامتن (HTML) از عناصر هسته دوبلین بیشتر است. و تکرار کلیدواژه‌ها از سوی موتورهای کاوش تقلب محسوب می‌شود.

پژوهش‌های رویکرد سوم بیشتر ماهیت مقایسه‌ای داشتند. برآیند این پژوهش‌ها، فرابرچسب‌های زبان نشانه‌گذاری فرامتن (HTML) را، همانند پژوهش‌های مرتبط با رویکردهای اول و دوم، اثر بخش‌تر از عناصر فراداده‌ای هسته دوبلین –به عنوان طرحی استاندارد– می‌دانست. به عبارت دیگر، نرم افزارهای

روبات موتورهای کاوش در هنگام پویش صفحاتی که هم از فرابرچسب‌ها و هم عناصر فراداده‌ای استفاده می‌کنند، فقط به فرابرچسب‌ها واکنش نشان می‌دهند. برخی دیگر از یافته‌ها، از کاهش تاثیر برخی ساز و کارهای مربوط به "بهینه‌سازی موتورهای کاوش (SEO)[1]" بر افزایش رتبه اشیای محتوایی در نتایج موتورهای کاوش خبر می‌داد.

محدودیت‌های زبان نشانه‌گذاری فرامتن (HTML) همچنان باقی بود. و موتورهای کاوش وب نیز حرکت جدی و مناسبی برای تغییر خط مشی‌های نمایه‌سازی و رتبه‌بندی خود در جهت بهبود فرایند نمایه‌سازی و پیدانمایی پیشینه‌های فراداده‌ای مبتنی بر طرح‌های استاندارد در بستر زبان نشانه‌گذاری فرامتن نشان نمی‌دادند.

فراهم کنندگان خدمات اطلاعاتی اصلی و معتبر وب (نظیر بافت کتابخانه‌ای و پایگاه‌های اطلاعاتی) با چالشی جدی در افزایش دسترس‌پذیری اشیای محتوایی خود از طریق موتورهای کاوش وب روبرو بودند. از سوی دیگر، طرح‌های فراداده‌ای که امکان پیاده‌سازی در بستر نحوی زبان نشانه‌گذاری فرامتن داشتند (مانند هسته دوبلین)، به دلیل نقاط ضعف مختلف، دوران انحطاط خود را می‌گذراندند. عامل دیگر، سازگاری طرح‌های فراداده‌ای پیشین نظیر مارک با محیط اطلاعاتی جدید بود که منشاء تولید طرح‌های فراداده‌ای کارآمدتر برای سازماندهی اشیای محتوایی دسترس‌پذیر در محیط شبکه‌ای شد. بدیهی است با این تحولات، فراهم کنندگان خدمات اطلاعاتی پیوسته به استفاده از طرح‌های جدید تمایل بیشتری نشان می‌دادند. اکنون این پرسش مطرح می‌شود که برای بهبود میانکنش‌پذیری موتورهای کاوش با نظام‌های فراداده‌ای استاندارد، طراحان و گسترش دهندگان نظام‌های فراداده‌ای باید چه تمهیداتی می‌اندیشیدند؟

تغییر بستر نحوی پیشینه‌های فراداده‌ای و واکنش مثبت موتورهای کاوش

گذار از بستر نحوی زبان نشانه‌گذاری فرامتن (HTML) به زبان نشانه‌گذاری گسترش‌پذیر (XML) برای پیاده‌سازی و ذخیره‌سازی پیشینه‌های فراداده‌ای، پاسخ گسترش دهندگان طرح‌های فراداده‌ای به چالش‌های میانکنش‌پذیری نظام‌های فراداده‌ای با موتورهای کاوش عمومی وب بود. زبان نشانه‌گذاری گسترش‌پذیر (XML) به عنوان استانداردی برای بازنمود و تبادل داده‌ها بر روی وب در سال ۱۹۹۸ توسط کنسرسیوم وب جهانی توسعه یافت. داده‌هایی که در قالب زبان XML نشانه‌گذاری می‌شوند،

[1]. Search Engine Optimization (SEO)

به داده‌هایی ساختارمند تبدیل، و مدارکی خود- توصیف[۱] بوجود می‌آورند. این ویژگی موجب استقلال اشیای محتوایی مبتنی بر زبان نشانه‌گذاری گسترش‌پذیر از هر سکویی[۲] شده، تبادل آنها را میان نظام‌های ناهمگن ممکن، و بنابراین میانکنش‌پذیری نظام‌های اطلاعاتی را تحقق می‌بخشد.

ساختار اشیای مبتنی بر XML، ساختاری سلسله مراتبی است. این ساختار موجب پردازش بهتر آنها شده، و از سوی دیگر، پیاده‌سازی و مستندسازی الگوها و فرایندهای بسیار پیچیده را میسر می‌سازد. قابلیت‌ها و ویژگی‌های منحصر بفرد این زبان، گرایش متخصصان و سازمان‌های فعال در حوزه مدیریت اطلاعات را، به پیاده سازی و دوباره قالب بندی استانداردها و طرح‌های این حوزه در قالب XML باعث گردیده است. تلاش‌های گسترده‌ای که در این حوزه انجام گرفته، دلیلی بر این مدعاست (طاهری و دیگران، ۱۳۸۸).

قابلیت‌های پیش‌گفته، تحقق آرمان‌های دیرین حوزه سازماندهی دانش را نوید می‌دهد. بنابراین امکان پیاده‌سازی همه انواع طرح‌های فراداده‌ای در بستر نحوی زبان XML وجود دارد.

در حال حاضر سه رویکرد کلی برای بازیابی اشیای محتوایی مبتنی بر زبان نشانه‌گذاری گسترش‌پذیر (XML) در نظام‌های بازیابی اطلاعات وجود دارد. رویکرد مبتنی بر پایگاه داده‌ها[۳]، رویکرد مبتنی بر بازیابی اطلاعات[۴]، و رویکرد دو وجهی[۵]. اشیای مبتنی بر زبان نشانه‌گذاری گسترش‌پذیر ، اشیای متنی با نشانه‌گذاری اضافی قمداد می‌شوند. برای پرداختن به برچسب‌های نشانه‌گذاری چندین روش وجود دارد. یک روش، حذف همه برچسب‌هاست. مزیت این روش، سادگی آن می‌باشد، و زیان آن از دست دادن اطلاعات و کاهش کارآیی بازیابی است. روش دیگر، استخراج اطلاعات ساختاری و بافتی از اشیای محتوایی (داده و فراداده) مبتنی بر XML برای نمایه‌سازی است. روش جامع‌تر، نمایه‌سازی برچسب‌ها به عنوان اصطلاحات نمایه‌ای است (لوک و دیگران، ۲۰۰۰).

جدیدترین پژوهش‌های حوزه میانکنش‌پذیری نظام‌های فراداده‌ای و موتورهای کاوش وب، به بررسی نمایه‌سازی و رتبه‌بندی پیشینه‌های فراداده‌ای مبتنی بر زبان نشانه‌گذاری گسترش‌پذیر (XML) پرداختند. بررسی تطبیقی نمایه‌سازی و رتبه‌بندی اشیای محتوایی حاوی عناصر فراداده‌ای هسته دوبلین و مارک ۲۱ توسط موتورهای کاوش عمومی، موضوع پژوهش طاهری (۱۳۸۷الف) بود. وی ۱۰۰ پیشینه

---

[1] . Self-description
[2] . platform
[3] . Database-based approach
[4] . Retrieval-based approach
[5] . Hybrid approach

فراداده‌ای مبتنی بر قالب فراداده‌ای مارک ۲۱ ، و طرح فراداده‌ای هسته دوبلین در بستر نحوی زبان نشانه‌گذاری گسترش‌پذیر را در کتاب‌های الکترونیکی مرتبط جاسازی نمود. اشیای تولید شده بر روی دو حوزه فرعی در محیطی ایستا منتشر، و سپس به موتورهای کاوش گوگل و یاهو معرفی شدند. نتایج پژوهش نشان داد، ارزش‌های همه عناصر فراداده‌ای مربوط به پیشینه‌های فراداده‌ای منتشر شده توسط موتورهای کاوش مورد مطالعه نمایه‌سازی شده‌اند. اما تفاوت معناداری میان رتبه‌بندی اشیای محتوایی مبتنی بر دو طرح مارک ۲۱ و هسته دوبلین نسبت به یکدیگر مشاهده نشد (طاهری ، ۱۳۸۷الف). بررسی زبان نشانه‌گذاری گسترش‌پذیر به عنوان بستر نحوی، و مطالعه پیشینه‌های فراداده مبتنی بر مارک ۲۱، نوآوری‌های پژوهش طاهری (۱۳۸۷الف) بود. پس از طاهری، طباطبایی و دیگران (۱۳۹۱)، و آقاعابدی (۱۳۸۹) بر اساس پیشنهادهای پژوهشی طاهری (۱۳۸۷الف)، دو پژوهش دیگر را طرح‌ریزی کردند. طباطبایی و دیگران (۱۳۹۱) جامعه پژوهشی شبیه پژوهش طاهری (۱۳۸۷الف) را در محیطی پویا[1] منتشر نموده، واکنش موتورهای کاوش به آنها را مورد ارزیابی قرار داد. یافته‌ها حاکی از آن بود، موتور کاوش گوگل، پیشینه‌های فراداده‌ای جاسازی شده در کتاب‌های الکترونی مبتنی بر زبان نشانه‌گذاری گسترش‌پذیر (XML) موجود در محیط‌های پویا را نیز نمایه‌سازی می‌کند، اما موتور کاوش یاهو عملکرد مثبتی نشان نداد. آقاعابدی (۱۳۸۹) رویکرد استفاده از فراداده را در پژوهش خود نسبت به طاهری (۱۳۸۷الف) و طباطبایی و دیگران (۱۳۹۱) تغییر داد. وی پیشینه‌های فراداده‌ای مبتنی بر زبان نشانه‌گذاری گسترش‌پذیر را مجزا بر روی وب سایت منتشر ساخت. نتایج پژوهش او با نتایج پژوهش طاهری (۱۳۸۷الف) یکسان بود. بدین معنی که همه موتورهای کاوش مورد مطالعه (گوگل و یاهو) نسبت به هر دو رویکرد استفاده از فراداده (جاسازی و پیوند) عملکرد مثبتی نشان دادند.

هر سه پژوهش مرتبط با رویکرد چهارم، اثربخشی بستر نحوی زبان نشانه‌گذاری گسترش‌پذیر (XML) در افزایش سطح بازیابی‌پذیری، و پیدانمایی پیشینه‌های فراداده‌ای طرح‌های استاندارد را تایید کردند. ویژگی‌های زبان نشانه‌گذاری گسترش‌پذیر (XML) موجب شد، هر سه پژوهش بتوانند دو طرح فراداده‌ای پرکاربرد را با یکدیگر مقایسه نمایند.

این امکان در پژوهش‌های رویکردهای پیشین عملی نبود. هر دو رویکرد استفاده از فراداده در سازماندهی اشیای محتوایی، و هر دو محیط ایستا و پویا مد نظر قرار گرفت. نکات قابل توجه درباره این سه پژوهش آن است که هر سه پژوهش در ایران انجام گرفته است،، با یکدیگر مرتبط بوده‌اند به شکلی

[1]. Dynamic environment

که ادامه یکدیگر قلمداد می‌شدند و جنبه‌های مختلف یک موضوع مورد پژوهش قرار گرفته است. به نظر می‌رسد یافته‌های این پژوهش‌ها از اعتبار مناسبی برخوردار باشند.

استفاده از زبان نشانه‌گذاری گسترش‌پذیر (XML) به عنوان بستر نحوی پیشینه‌های فراداده‌ای، میانکنش‌پذیری نظام‌های فراداده‌ای و موتورهای کاوش عمومی وب را افزایش می‌دهد (طاهری و دیگران، ۱۳۸۸). بنابراین توجه به این مهم، افزایش سطح دسترس‌پذیری اشیای محتوایی از طریق ابزارهای پرکاربرد وب را در پی خواهد داشت. با این وجود، موتورهای کاوش برای نمایه‌سازی اشیای محتوایی در قالب زبان نشانه‌گذاری گسترش‌پذیر رویکرد حذف برچسب را برگزیده‌اند (طاهری و حریری، ۲۰۱۲). به عبارت دیگر، خود برچسب‌ها به عنوان اصطلاحات نمایه‌ای نمایه‌سازی نمی‌شوند، و فقط ارزش‌های آنها استخراج می‌شود. دلیل انتخاب این رویکرد از سوی موتورهای کاوش، گسترش‌پذیر بودن زبان XML است. یعنی برچسب‌های از پیش تعریف شده‌ای در این زبان وجود ندارد تا به نرم افزارهای خزنده- نمایه‌ساز موتورهای کاوش از پیش معرفی شوند. بلکه هر پدیدآورنده حقیقی یا حقوقی در زمان ایجاد اشیای محتوایی مبتنی بر آن زبان، آنها را طراحی می‌کند.

این مسئله، بهره‌مندی از ساختار سلسله مراتبی و منعطف را به حداقل می‌رساند. به بیان ساده، کمترین پردازش بر اشیای محتوایی مبتنی بر زبان نشانه‌گذاری گسترش‌پذیر (XML) از سوی موتورهای کاوش وب انجام می‌پذیرد. بنابراین کارایی فرایند بازیابی در این موتورها کاهش می‌یابد. راهکار حوزه سازماندهی دانش برای عبور از این مشکل چیست؟

نتیجه‌گیری

همان طور که پیشتر گفته شد، ناکارآمدی بستر نحوی زبان نشانه‌گذاری فرامتن (HTML) برای ذخیره‌سازی عناصر طرح‌های فراداده‌ای استاندارد به دلیل محدودیت تعداد برچسب‌ها و فرابرچسب‌ها، کارکرد صرفا نمایشی و واکنش منفی نرم افزارهای خزنده- نمایه ساز موتورهای کاوش به فرابرچسب‌هایی غیر فرابرچسب‌های آن زبان، موجب گرایش گسترش دهندگان طرح‌های فراداده‌ای به تغییر بستر نحوی فراداده‌ها شد. جایگزینی زبان نشانه‌گذاری گسترش‌پذیر (XML) به عنوان بستر نحوی اقدامی راهبردی بود که محدودیت‌های بستر پیشین را بر طرف نمود.

بستر جدید علاوه بر امکان ذخیره‌سازی و پیاده‌سازی انواع طرح‌های فراداده‌ای، افزایش میانکنش‌پذیری نظام‌های فراداده و موتورهای کاوش را در پی داشت. بنابراین امکان مقایسه طرح‌های استاندارد نیز به وجود آمد. قابلیت تولید پیشینه‌های فراداده‌ای بر اساس رویکرد پیوند، و برقراری ارتباطات

معنادار و نظام‌مند با اشیای محتوایی مرتبط را فراهم نمود، که به نو به خود، باعث افزایش کاربردپذیری زبان نشانه‌گذاری گسترش‌پذیر (XML) در محیط‌های اطلاعاتی کتابخانه‌ای -بویژه فهرست‌های رایانه‌ای- و پایگاه‌های اطلاعات علمی معتبر می‌شود.

اشیای محتوایی (داده و فراداده) مبتنی بر زبان نشانه‌گذاری گسترش‌پذیر (XML) را می‌توان هم در محیط‌های ایستا و هم محیط‌های پویا منتشر نمود. این ویژگی زبان نشانه‌گذاری گسترش‌پذیر (XML) نیز برای بافت فراداده، ارزش افزوده تولید می‌کند.

از سوی دیگر، زمینه برای بهبود فرایند بازیابی در محیط موتورهای کاوش فراهم شده است. پیشینه‌های فراداده‌ای مستند- مانند طرح فراداده‌ای توصیف مستند(MADS)- را می‌توان درون اشیای محتوایی (داده یا فراداده) مرتبط جاسازی نمود، و بدین وسیله جامعیت نظام بازیابی موتورهای کاوش را افزایش داد.

با وجود بهبود میانکنش‌پذیری دو نظام مورد مطالعه با بهره‌مندی از زبان نشانه‌گذاری گسترش‌پذیر (XML)، رویکرد موتورهای کاوش در نمایه‌سازی اشیای محتوایی مبتنی بر XML، بافت فراداده را با چالشی دیگر روبرو کرده است. اگر چه پیشینه‌های فراداده‌ای مبتنی بر زبان نشانه‌گذاری گسترش‌پذیر (XML) توسط موتورهای کاوش نمایه‌سازی شده، قابل بازیابی هستند، اما سیاست حذف برچسب از کارایی بازیابی کاسته است. اندیشیدن و طراحی راهکارهایی برای حل این مشکل پژوهش‌هایی جدید را می‌طلبد. اگر موتورهای کاوش برچسب‌ها را نیز همانند ارزش‌های آنها نمایه‌سازی کنند، امکان جستجوهای عنصر- پایه[1] در نظام جستجوی آنها به وجود می‌آید.

این ویژگی برای هیچ یک از قالب‌های ذخیره سازی (HTML، PDF، WORD، و ...) میسر نیست. انتظار می‌رود چنین راهکارهایی افقی جدید را در دسترس‌پذیری اشیای محتوایی وب بگشاید. جدول شماره ۱ رویکردهای بیان شده در مقاله و تلاش‌های مرتبط را به صورت مختصر نشان می‌دهد:

جدول شماره ۱. رویکردهای حوزه میانکنش‌پذیری و تاثیر آنها بر سازماندهی اشیای محتوایی وب

رویکردهای حوزه میانکنش‌پذیری	تاثیر آنها بر سازماندهی اشیای محتوایی وب و واکنش موتورهای کاوش
برچسب‌ها و فرا برچسب‌های زبان نشانه گذاری فرامتن (HTML) و موتورهای کاوش	تکیه بر برچسب‌ها و فرابرچسب‌های HTML به عنوان عناصر فراداده‌ای برای توصیف اشیا (صفحات)؛ واکنش مثبت موتورهای کاوش به نمایه سازی و پیدانمایی آنها؛ عدم استقبال محیط‌های اطلاعاتی کتابخانه‌ای و پایگاه‌های تمام متن به دلیل محدودیت‌های زبان HTML

[1]. Element-base

	(معدود بودن برچسب‌ها و فرابرچسب‌ها؛ کارکرد صرفا نمایشی آن زبان؛ و ...)
پیشینه‌های فراداده‌ای مبتنی بر طرح‌های فراداده‌ای استاندارد و موتورهای کاوش	جاسازی پیشینه‌های فراداده‌ای برخی از طرح‌های فراداده‌ای در زبان HTML؛ واکنش منفی موتورهای کاوش به آن پیشینه‌ها؛ و در نتیجه عدم استقبال محیط‌های اطلاعاتی کتابخانه‌ای و پایگاه‌های تمام متن
رویکرد دو وجهی (فرا برچسب‌های HTML و استانداردهای فراداده‌ای)	تولید پیشینه‌های فراداده بر اساس برخی استانداردهای فراداده‌ای و برچسب‌ها و فرابرچسب‌های HTML به صورت ترکیبی؛ واکنش منفی موتورهای کاوش به پیشینه‌های استاندارد (اما مثبت به برچسب‌ها و فرابرچسب‌های HTML)؛ انجام پژوهش‌های مقایسه‌ای؛ عدم استقبال محیط‌های اطلاعاتی کتابخانه‌ای و پایگاه‌های تمام متن
تغییر بستر نحوی پیشینه‌های فراداده‌ای، و واکنش مثبت موتورهای کاوش	استفاده از زبان نشانه گذاری گسترش پذیر (XML) به عنوان بستر نحوی پیشینه‌های فراداده‌ای؛ حذف محدودیت‌های زبان HTML؛ واکنش کاملا مثبت موتورهای کاوش به آن پیشینه‌ها؛ بنابراین استقبال محیط‌های اطلاعاتی کتابخانه‌ای و پایگاه‌های تمام متن

کتابنامه

آقاعابدی، زهرا (۱۳۸۹). بررسی تطبیقی کیفیت نمایه سازی و رتبه بندی پیشینه‌های فراداده‌ای هسته دوبلین و مارک ۲۱ توسط موتورهای کاوش عمومی . پایان نامه کارشناسی ارشد. رشته کتابداری و اطلاع رسانی. دانشگاه آزاد اسلامی، واحد علوم و تحقیقات تهران.

شریف، عاطفه (۱۳۸۶). "بررسی میزان اثر بخشی عناصر ابرداده‌ای بر رتبه بندی صفحات وب توسط موتورهای کاوش". فصلنامه کتابداری و اطلاع رسانی. شماره ۳۸ (تابستان ۱۳۸۶) [پیوسته]، دسترس پذیر: http://aqlibrary.ir/index.php?module=TWArticles&file=index&func=view_pubarticles&did=145&pid=10 . [۳۰ آبان ۱۳۹۰].

طاهری، مهدی (۱۳۸۷الف). *"بررسی تطبیقی کیفیت نمایه سازی و رتبه بندی اشیای محتوایی حاوی عناصر فراداده‌ای هسته دوبلین و مارک ۲۱ توسط موتورهای کاوش عمومی"*. پایان نامه کارشناسی ارشد. پایان نامه کارشناسی ارشد. رشته کتابداری و اطلاع رسانی. دانشگاه آزاد اسلامی، واحد علوم و تحقیقات تهران.

طاهری، مهدی (۱۳۸۷ب). "مقایسه کارایی طرح فراداده‌ای هسته دوبلین و قالب فراداده‌ای مارک ۲۱ در سازماندهی منابع اطلاعاتی شبکه جهانی وب". *فصلنامه کتابداری و اطلاع رسانی آستان قدس رضوی*. شماره ۴۳ (پاییز ۱۳۸۷).۶۶-۷۹.

طاهری، مهدی؛ حریری، نجلا؛ فتاحی، رحمت‌الله (۱۳۸۸). "بررسی تطبیقی کیفیت نمایه سازی و رتبه بندی اشیای محتوایی حاوی عناصر فراداده‌ای هسته دوبلین و مارک ۲۱ توسط موتورهای کاوش عمومی". *فصلنامه کتابداری و اطلاع رسانی آستان قدس رضوی*. شماره ۴۸ (زمستان ۱۳۸۸).۱۴۱-۱۶۲

طباطبایی امیری، فائزه السادات؛ طاهری، سید مهدی؛ فرج پهلو، حسین؛ عصاره، فریده؛ معرف‌زاده، عبدالحسین (۱۳۹۱). "موتورهای کاوش وب، و نمایه‌سازی و رتبه‌بندی اشیای محتوایی حاوی عناصر فراداده‌ای در محیط‌های پویا ". *فصلنامه پردازش و مدیریت اطلاعات*. دوره ۲۷، شماره ۴ (تابستان ۱۳۹۱). ۹۰۷-۹۲۰.

فتاحی، رحمت‌الله؛ طاهری، مهدی؛ (۱۳۸۸).*" فهرستنویسی رایانه‌ای: مفاهیم، شیوه‌ها، و کاربرد نرم افزارهای رایانه-ای در سازماندهی اطلاعات"*. با همکاری فرشته ناقدی احمدی. تهران: کتابدار

Bifet, Albert; Castillo, Carlos (2005). "An Analysis of Factors Used in Search Engine Ranking". [online], available at: http://airweb.cse.lehigh.edu/2005/bifet.pdf. [14 Dec. 2011].

Google (2012). "Web Master Tools: Meta tag". [online], available at http://www.google.com/support/webmasters/bin/answer.py?answer=79812. [5 Jun. 2011].

Henshaw, Robin; Valauskas, Edward J (2001). " Metadata as a Catalyst: Experiments with Metadata and Search Engines in the Internet Journal, First Monday ". [online], available at: www.librijournal.org/pdf/1999-3pp125-131.pdf . [5 Dec. 2010].

Luk, Robert, et al (2000). "A Survey of Search Engines for XML Documents".[online],

available at: http://www.haifa.il.ibm.com/sigir00-xml/final-papers/Luk/XMLSUR.htm. [14 Dec. 2011].

Mohamed, Khaled A.f. 2006. The impact of metadata in web resources discovering. Online Information Review. 30 (2): 155-167

Quevedo-Torrero, Jesus Ubaldo (2004). "IMPROVING WEB RETRIEVAL BY MINING THE HTML TAGS FOR KEYWORDS AND EXPLORING THE HYPERLINK STRUCTURES WEB PAGES". Ph. D. Dissertation, Department of Computer Science, University of Houston. [online], available at: http://wwwlib.umi.com/dissertations/fullcit/3156028. [5 Dec. 2011].

Safari, Mehdi (2005). "Search Engine and Resource Discovery on the Web: Is Dublin Core an Impact Factor". [online], available at: www.webology.ir/2005/v2n2/a13.html . [5 Dec. 2011].

Search Engine Tutorial (2012). "Metatags". [online], available at http://www.searchenginetutorial.com/metatags.html . [5 Jun. 2011].

Sokvitne, Lloyd (2000). "An Evaluation of the Effectiveness of Current Dublin Core Metadata for Retrieval". [online], available at: www.vala.org.au/vala2000/2000pdf/Sokvitne.PDF . [7 Dec. 2011].

Taheri, S. M ;.Hariri, Nadjla (2012). "A Comparative Study on the Indexing and Ranking of the Content Objects Including the MARCXML and Dublin Core's Metadata Elements by General Search Engines". Electronic Library (ISI ranked journal). Vol 30, issue 4

Turner, Thomas P.; Brackbill, Lise (1998). "Rising to the Top: Evaluating the Use of the HTML META Tag To Improve Retrieval of World Wide Web Documents through Internet Search Engines". [online], available at: http://cat.inist.fr/?aModele=afficheN&cpsidt=1748620 . [7 Dec. 2011].

Yahoo (2012). "HELP: What are meta tags?". [online], available at: http://help.yahoo.com/l/us/yahoo/smallbusiness/promotion/meta/meta-01.html. [5 Jun. 2011].

Zhang, Jin; Dimitroff, Alexandra (2004). "Internet search engine's response to metadata Dublin Core implementation ". [online], available at: http://portal.acm.org/citation.cfm?id=1142111 . [5 Dec. 2011].

Zhang, Jin; Dimitroff, Alexandra (2005a). "The impact of Webpage content characteristics on webpage visibility in search engine result (Part I)". [online], available at: http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6VC8-4BHCBX4-1&_user=10&_rdoc=1&_fmt=&_orig=search&_sort=d&view=c&_acct=C000050221&_version=1&_urlVersion=0&_userid=10&md5=33927751b92200b392f8c79b950dcdb1 . [14 Dec. 2011].

Zhang, Jin; Dimitroff, Alexandra (2005b). "The impact of metadata implementation on Webpage visibility in search engine result (Part II)". [online], available at: http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6VC8-4BHCBX4-2&_user=10&_rdoc=1&_fmt=&_orig=search&_sort=d&view=c&_acct=C000050221&_version=1&_urlVersion=0&_userid=10&md5=a853d410a866732d3f8ab5dd3217d412 . [5 Dec. 2011].